

Apache Flink Word Count App

User

2025-01-17, 06:50:06

Generate a Apache Flink application using Scala language to solve the problem! Given a plain text UTF-8 file URL and a output CSV file URL, write in the output file the occurrence sum of each word in the input file.

Assistant Phi 3 mini Instruct

2025-01-17, 06:50:06

Here is a Scala-based Apache Flink application that reads a plain text UTF-8 file from a given URL, counts the occurrences of each word, and writes the results to a CSV file at a specified output URL. ```scala import org.apache.flink.api.java.ExecutionEnvironment import org.apache.flink.api.scala.DataSet import org.apache.flink.api.scala.tuple.Tuple2 import org.apache.flink.streaming.api.scala.StreamExecutionEnvironment import org.apache.flink.streaming.api.scala.input.FileInputFormat import org.apache.flink.streaming.api.scala.output.TextOutput import org.apache.flink.streaming.api.scala.{StreamFunction, StreamExecutionEnvironment} import org.apache.flink.streaming.api.windowing.time.Time import org.apache.flink.streaming.api.functions.sink.FileSink import org.apache.flink.streaming.api.windowing.windows.TimeWindow import org.apache.flink.streaming.api.windowing.assigners.CountWindowAssigner import org.apache.flink.streaming.api.windowing.windows.GlobalWindows import org.apache.flink.streaming.api.functions.windowing.WindowFunction import org.apache.flink.streaming.api.scala.StreamExecutionEnvironment import org.apache.flink.streaming.api.scala.DataStream import org.apache.flink.streaming.api.scala.functions.RichWindowFunction import org.apache.flink.streaming.api.scala.functions.RichWindowFunction object WordCountApp extends App { val env = StreamExecutionEnvironment.getExecutionEnvironment // Read the input file val inputFile = "input-file-url" val text = env.readTextFile(inputFile) // Split the text into words and count occurrences val words = text .flatMap(_.toLowerCase.split("\\W+")) .filter(_.nonEmpty) .map(_._1) // Group by word and count occurrences val wordCounts = words .keyBy(0) .sum(1) // Write the results to the output CSV file val outputFile = "output-csv-file-url" wordCounts.writeAsText(outputFile) env.execute("WordCountApp") } ``` Please replace ``input-file-url`` with the actual URL of the input text file and ``output-csv-file-url`` with the desired output CSV file URL. This application uses Apache Flink's streaming API to process the input text file. It reads the file, splits the text into words, counts the occurrences of each word, and writes the results to a CSV file. Note that this example assumes that the input file is small enough to fit into memory. For larger files, you may need to consider using a distributed file system or a different approach for reading and processing the data. Before running this application, make sure you have the necessary dependencies for Apache Flink and Scala in your project. You can find more information on how to set up and run Apache Flink applications in the official documentation: <https://ci.apache.org/projects/flink/flink-docs-stable/dev/streaming/quickstart.html><|end|>